IMPORTANT: Infiles folder and Outfiles folder contain Jill’s Master Files to help us save time as running some of the analyses would be too long. Do not save any results to these folders. All of your work will be saved in Student\_folder.

**Morbilliviruses and host receptors: A case study with Nectin4 (poliovirus receptor-related protein 4 PVRL4) and Canine Distemper Virus**

Background:

There are seven known Morbilliviruses that infect Carnivores, Cetaceans, Primates, ungulates, small ruminants:

Feline morbilliviruses

Canine distemper virus

Phocine morbillivirus

Rinderpest virus

Small ruminant morbillivirus

Measles morbillivirus

Cetacean Morbillivirus

CDV is known to access cell entry into the host through host genes SLAM and NECTIN4 (PVRL4). It is hypothesized that NECTIN4 is linked with neurological forms of canine distemper. We will conduct a comparative genomic analysis of NECTIN4 in mammals to determine structure, function and evolution. Our study will focus only on the coding regions (CDS) of the gene.

We have searched RefSeq and NCBI for CDS (exons) of Nectin4 from Primates, Artiodactyls, Carnivores and Cetaceans and downloaded representative sequences that are full-length CDS of NECTIN4.

**PART 1: ALIVIEW-program for alignment and editing.**

**Open ALIVIEW in APPS folder**

**Step 1: Import Sequences in Aliview for Alignment with Muscle.**

**File-Open ‘***Desktop*/*Alignment\_Exercise /infiles/‘32 documents from GDW2018\_Nectin4\_Data’*

**Step 2: Perform Alignment**

**Selection-Select all**

**Align-Align everything as translated amino acids**

**Step 3: Trim alignment.** 5’ region Ma’s night monkey is added sequence that is a transcript variant. The same is true for the 3’ regions for Cat1 and Hawaiian monk seal.

Highlight positions 1-393 for all sequences

**Edit-delete selected**

Highlight positions 1555-1597 (terminal position) for all sequences

**Edit-delete selected**

**Save the edited file**

**File- Save Fasta file**  in  **‘***Desktop/Alignment\_Exercise/Student\_files /Nectin4\_trimmed.fasta’*

Remove the terminal stop codon TGA for all sequences. This is for downstream applications. Highlight positions 1552-1554.

**Edit-delete selected**

**ReSave the edited file**

**File- Save Fasta file**  in  **‘***Desktop/Alignment\_Exercise/Student\_files /Nectin4\_trimmed.fasta’*

**Step 4: View translation as codons**

**View-show as translation**

Codon triplets are now colored. You can visually inspect, particularly around indels, to be sure there are no false frame shift mutations or stop codon.

**Step 5: Local re-alignment**. Highlight region 37-73 for block realignment around indel.

**Align-Realign selected block.**

ReSave the edited file

**File- Save Fasta file**  in  **‘***Desktop/Alignment\_Exercise/Student\_files/Nectin4\_trimmed.fasta’*

**Step 6: Rename sequences as names are too long for subsequent programs.** To do so, just double-click on the name and enter it by hand. However, to save time I have already done this for you. To look at file

**File-Open** ‘*Desktop/Alignment\_Exercise/Outfiles/Nectin32TaxaMuscle.fasta’*

**Step 7: We can also combine two sets of sequences for alignment based on taxonomic lineages**.

**File-Open** ‘*Desktop/Alignment\_Exercise/Infiles/Nectin4primates.fasta’*

Align these sequences:

**Selection-select all**

**Align-Align all as translated amino acids**

Save

**File-Save as Fasta** ‘*Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesMuscle.fasta’*

Keeping *Nectin4PrimatesMuscle.fasta* open, we add the second set of sequences

**File Open** *Desktop/Alignment\_Exercise/Infiles/Nectin4Carnivores.fasta’*

**Selection-select all**

**Align-Align all as translated amino acids**

Save

**File-Save as Fasta** ‘*Desktop/Alignment\_Exercise/Student\_files/Nectin4CarnivoresMuscle.fasta’*

We can combine these two alignments by selecting all sequences in *Nectin4CarnivoresMuscle.fasta’*

copy

**Align**

**Add and Align sequences from clipboard**

We can then check this combined alignment to see if still consistent with codon structure.

**Selection-Select all**

**View-Show as translation**

Save & rename our results:

**Save-save as fasta** ‘*Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresMuscle.fasta’*

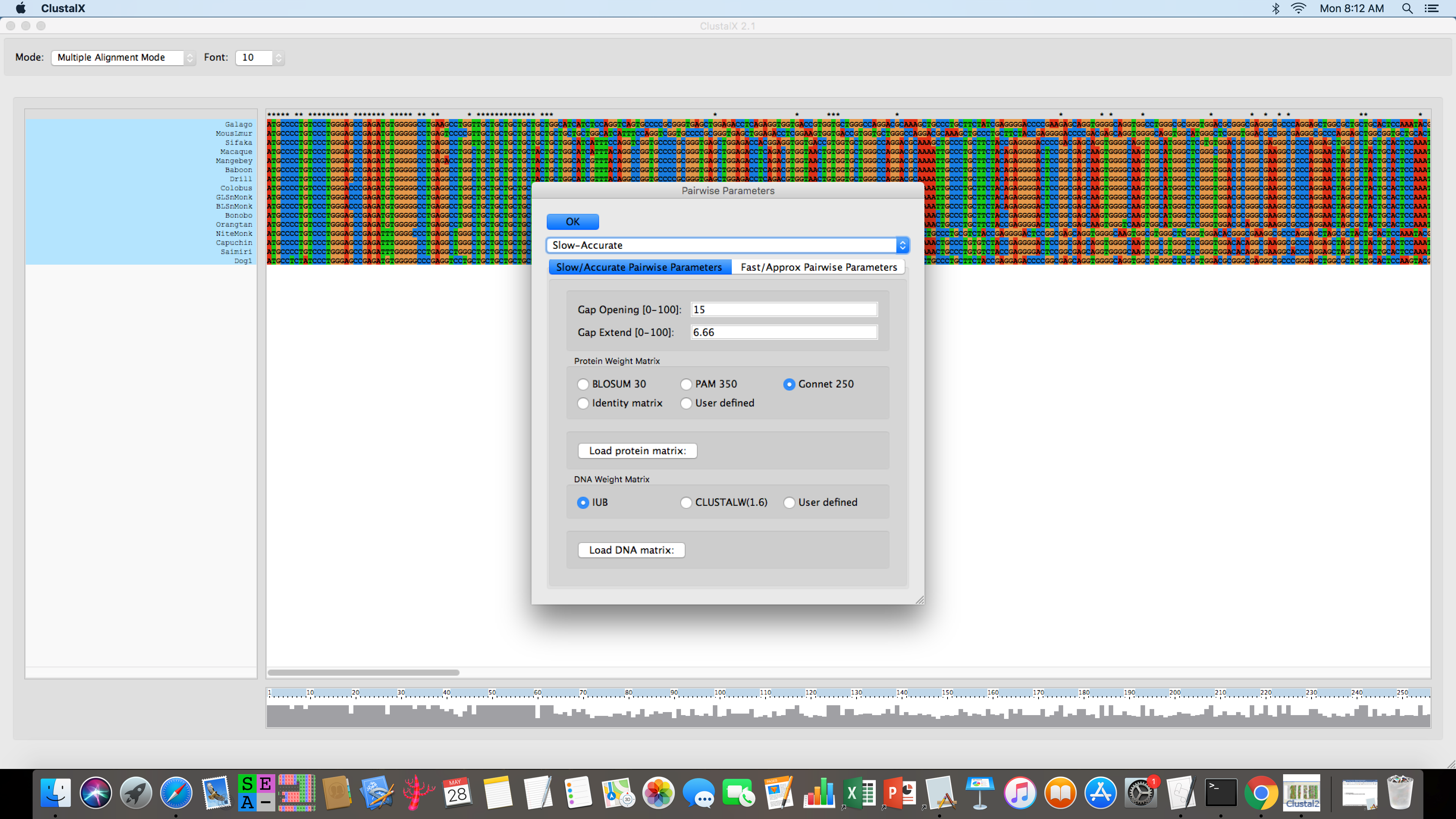
**Part 2: ClustalX Alignment using Lineage specific subsets Carnivore, Cetacean, Primates, and Ruminants**

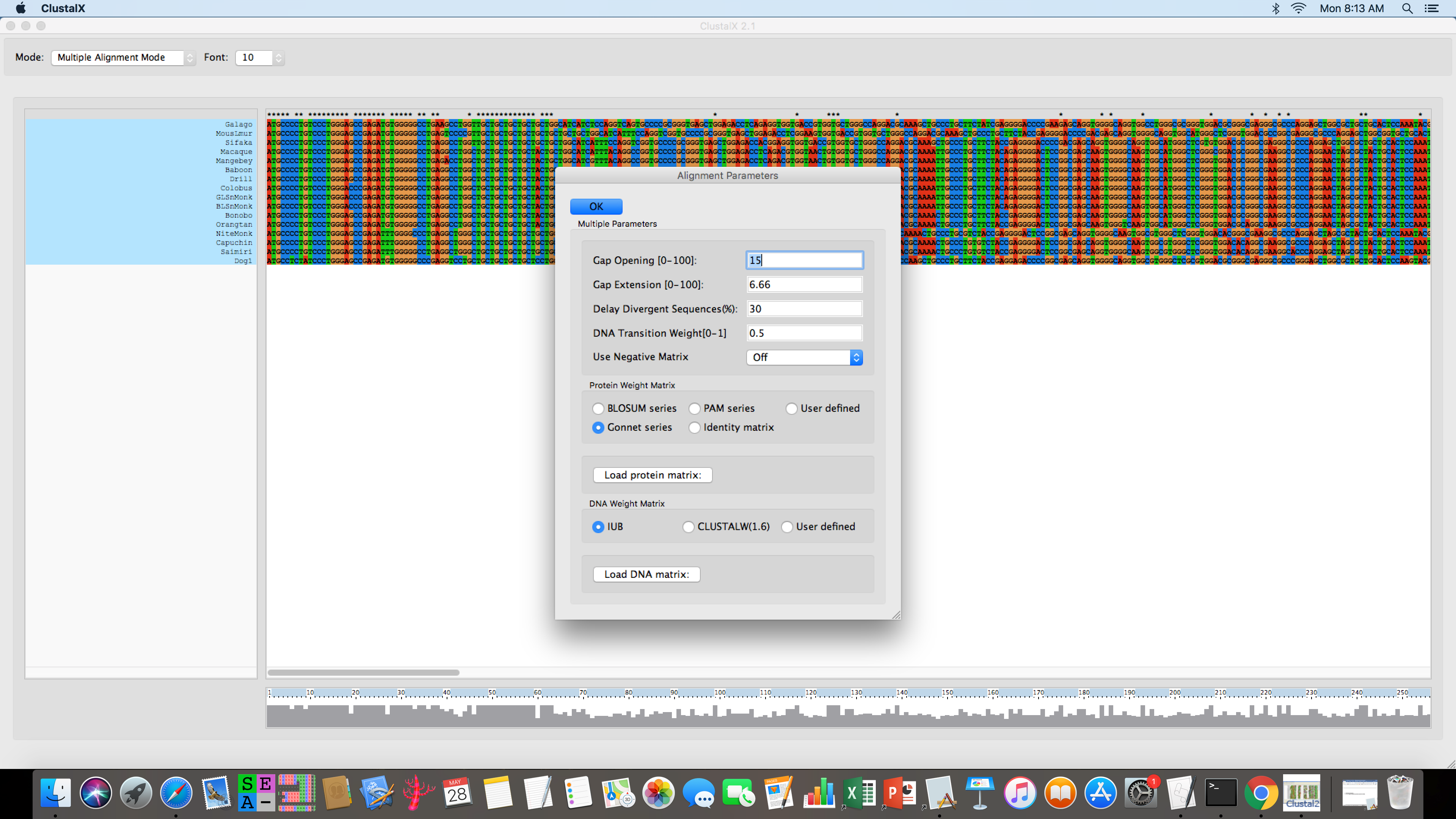
ClustalX has a useful tool in that we can switch between multiple alignment and profile alignment to help with step-wise addition of divergent sequences. It uses a different method for alignment than MUSCLE (Aliview).

We will use Clustal X because of the GUI interface, but the most recent version is more powerful & scalable Clustal Omega command line for those of you wanting to use this version. We will use both.

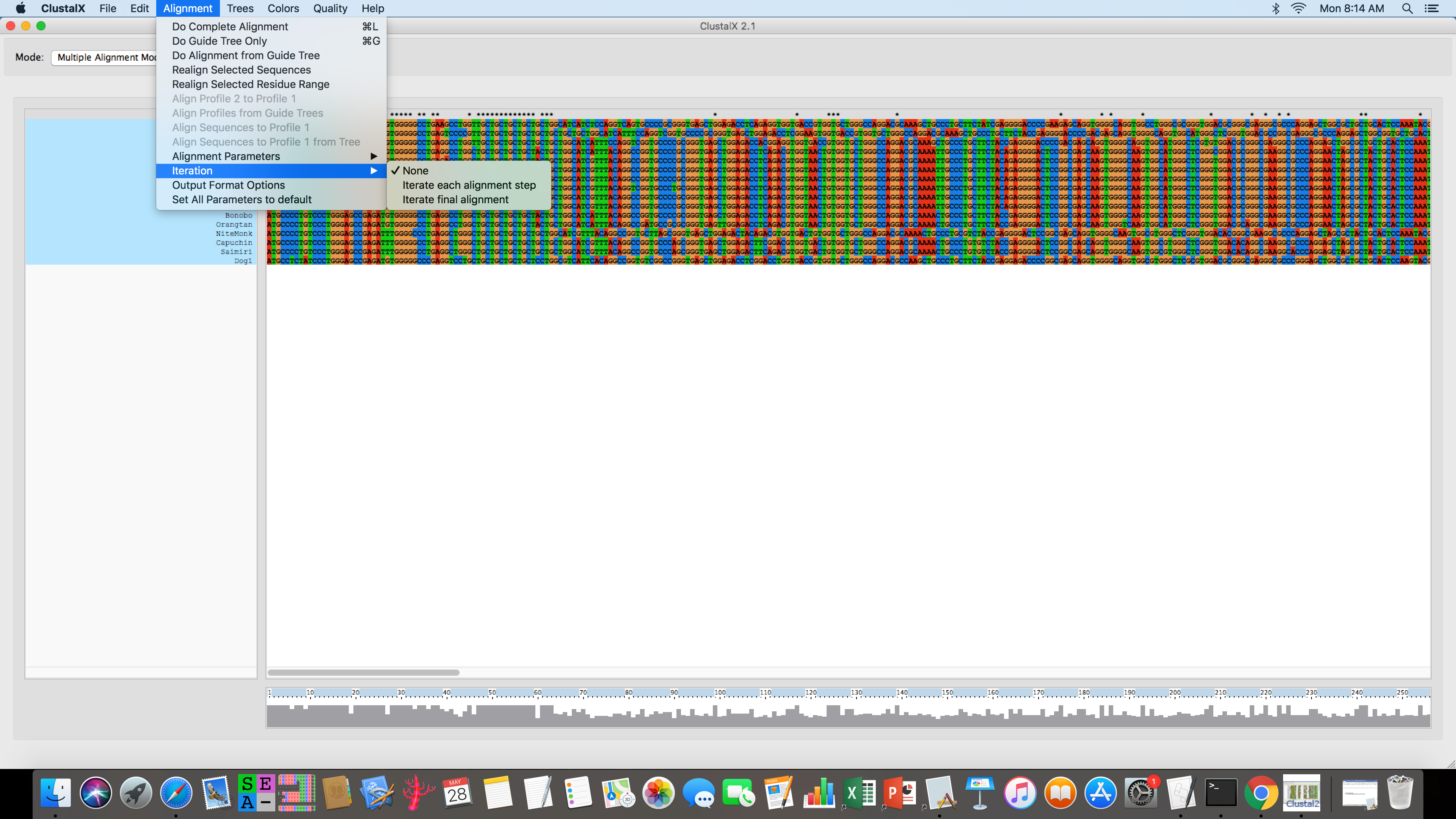
Website: http://www.clustal.org/clustal2/

1. **File-load sequences ‘***Desktop/Alignment\_Exercise/Clustal Folder/Nectin4PrimatesClus.fasta’*
2. **Edit-Select all sequences**
3. Examine **Alignment-alignment parameters-pairwise & multiple & Iteration.** These are very helpful with difficult alignments (divergent sequences), but we will go with default values for now.

****

****

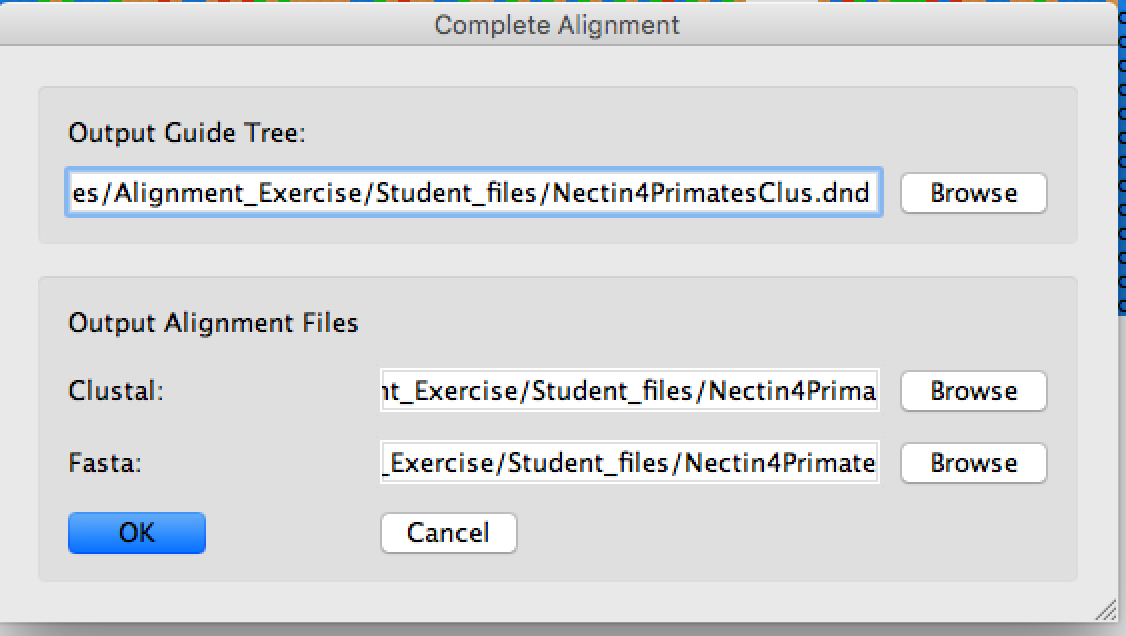
**The iteration function is useful, it will break apart and retest guide trees during each step of alignment or in the final alignment. It may help with more divergent homologous sequences.**

****

1. **Alignment-Do complete alignment**
   1. To avoid over-writing files, edit the path for saving the two outputs \*aln and \*.dnd to

*Desktop/Alignment\_Exercise /Student\_files /Nectin4PrimatesClus.aln*

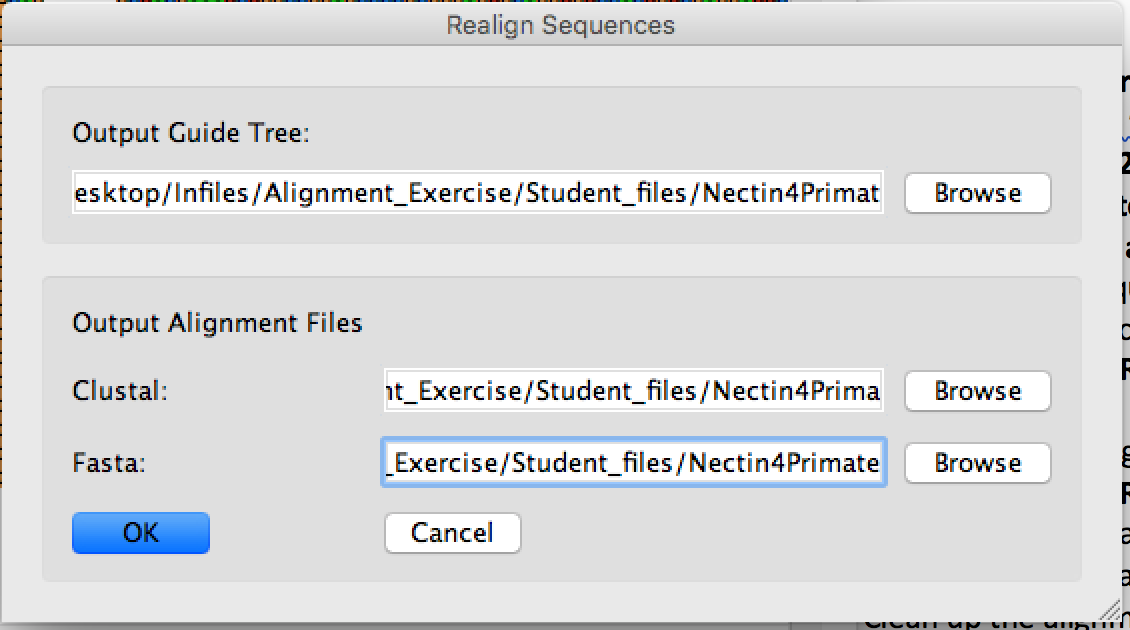
*Desktop/Alignment\_Exercise /Student\_files /Nectin4PrimatesClus.dnd*

****

1. **Change Mode to Profile alignment mode**
2. **File-Load Profile 2 ‘***Desktop/Alignment\_Exercise/Clustal/Nectin4CarnivoresClus.fasta’*
3. **Edit-Select Profile 2**
4. **Edit-Add Profile 2 to Profile 1**
5. **Switch to multiple alignment mode**
6. Realign specific sequences that look out of alignment (1.e. all the added Carnivore sequences) by highlighting specific sequences then:

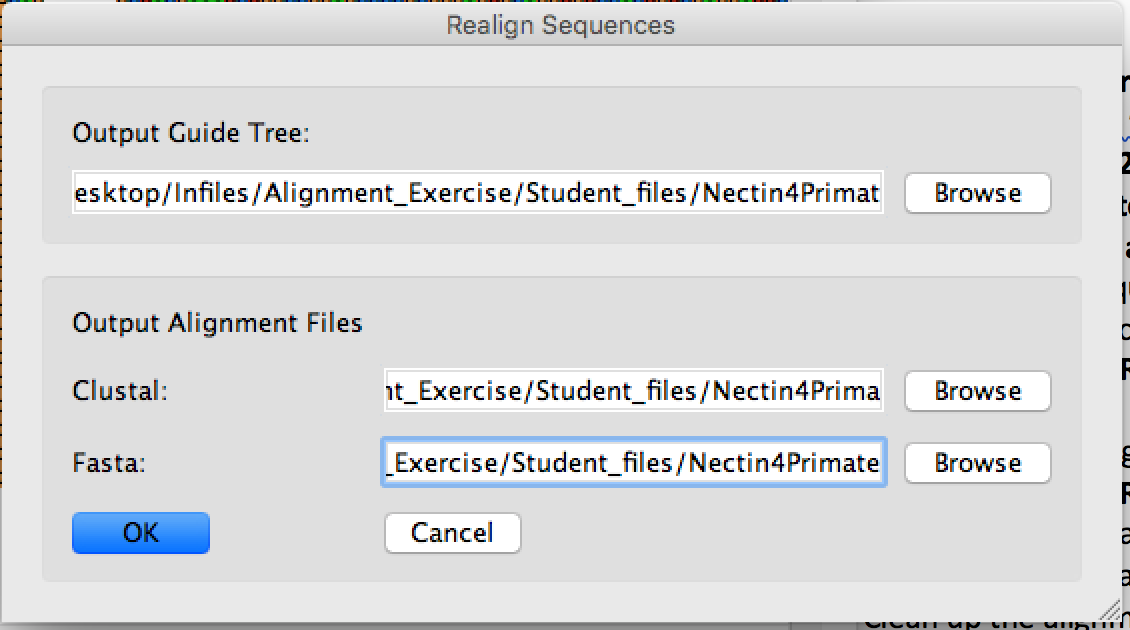
**Alignment-Realign selected sequence**

**Rename your files for both \*.aln and \*.dnd** *Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresClus*

****

1. Optional: realign selected regions around indel for re-alignment by highlighting block then:

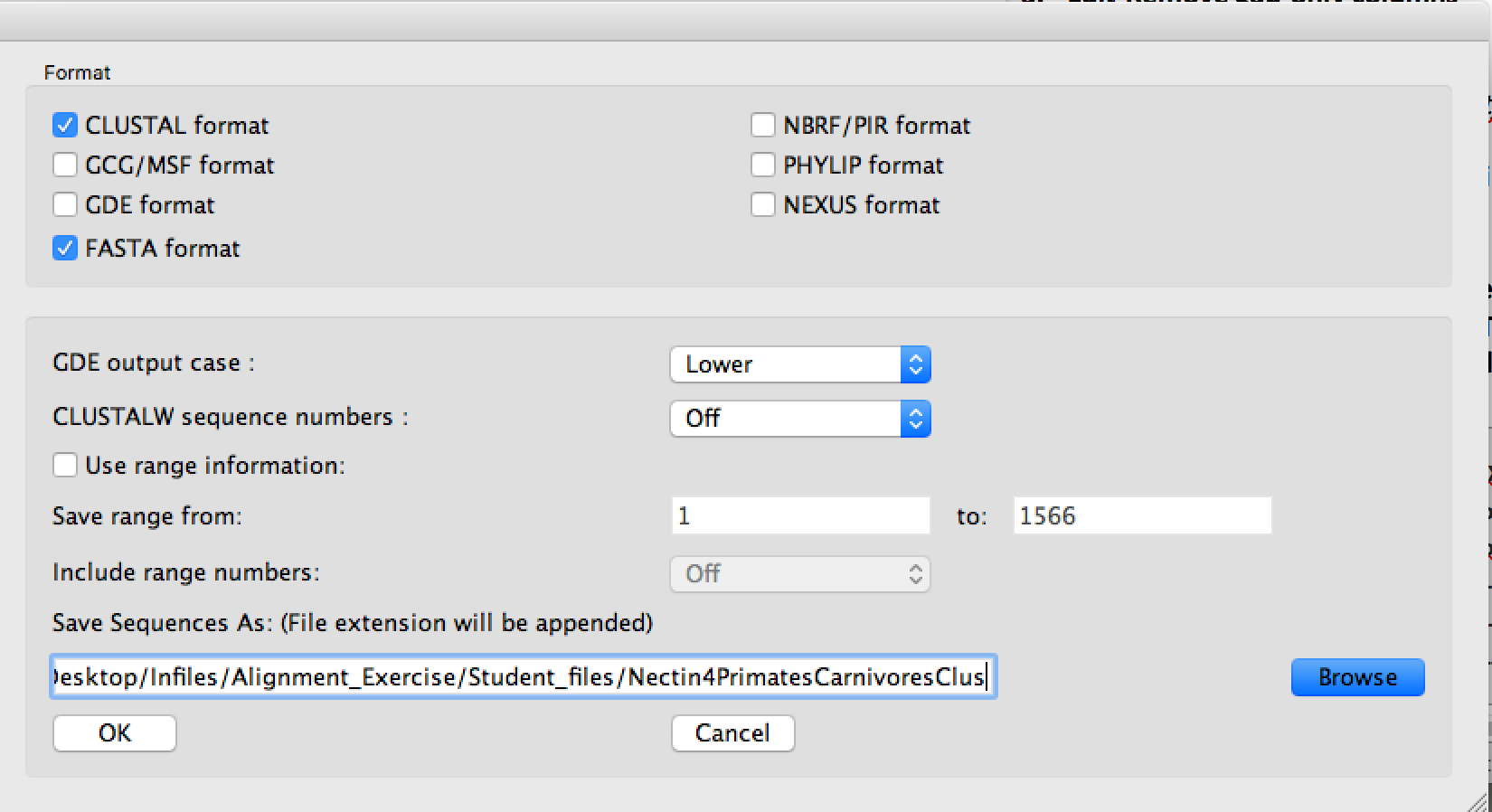
**Alignment-Realign selected residue range**

****

1. On your own, repeat this with changing parameters for pairwise, multiple and iteration to see if the indel region changes. Remember to keep saving your files into **Student\_files folder**
2. Clean up the alignment by removing gap-only columns if needed.
   1. **Edit-Remove gap-only columns**
3. Save final alignment file in both ClustalX and Fasta format
   1. **File-Save Sequences as**

*Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresClustal.fasta*

*Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresClustal.aln*

**

1. **Open this file in Aliview, and using the directions above for Part 1, view as codons. Do ClustalX and Muscle share same alignment? Open both files in Aliview to compare.**

**Using Clustal Omega for the same exercise.** (In the future, if you want to change some of the settings, type clustalo --help)

**Step 1: Open terminal and go to Clustal folder**

cd /Users/gdw/Desktop/Alignment\_Exercise/Clustal

**Step 2: Align Nectin4Primates.fasta and save it an outfile using default settings.**

clustalo -i Nectin4PrimatesClus.fasta -o Nectin4PrimatesOmega.out

**Step 3**: Open file in **Aliview** to check alignment

Step 4: Return to terminal and Clustal folder

cd /Users/gdw/Desktop/Alignment\_Exercise/Clustal

Step 5: Align Nectin4Carnivores and save to an outfile

clustalo -i Nectin4CarnivoresClus.fasta -o Nectin4CarnivoresOmega.out

**Step 6.** Open file in **Aliview** to check alignment

**Step 7:** Return to Terminal and Clustal folder.

cd /Users/gdw/Desktop/Alignment\_Exercise/Clustal

**Step 8**: Merge two profiles into single alignment and save to an outfile

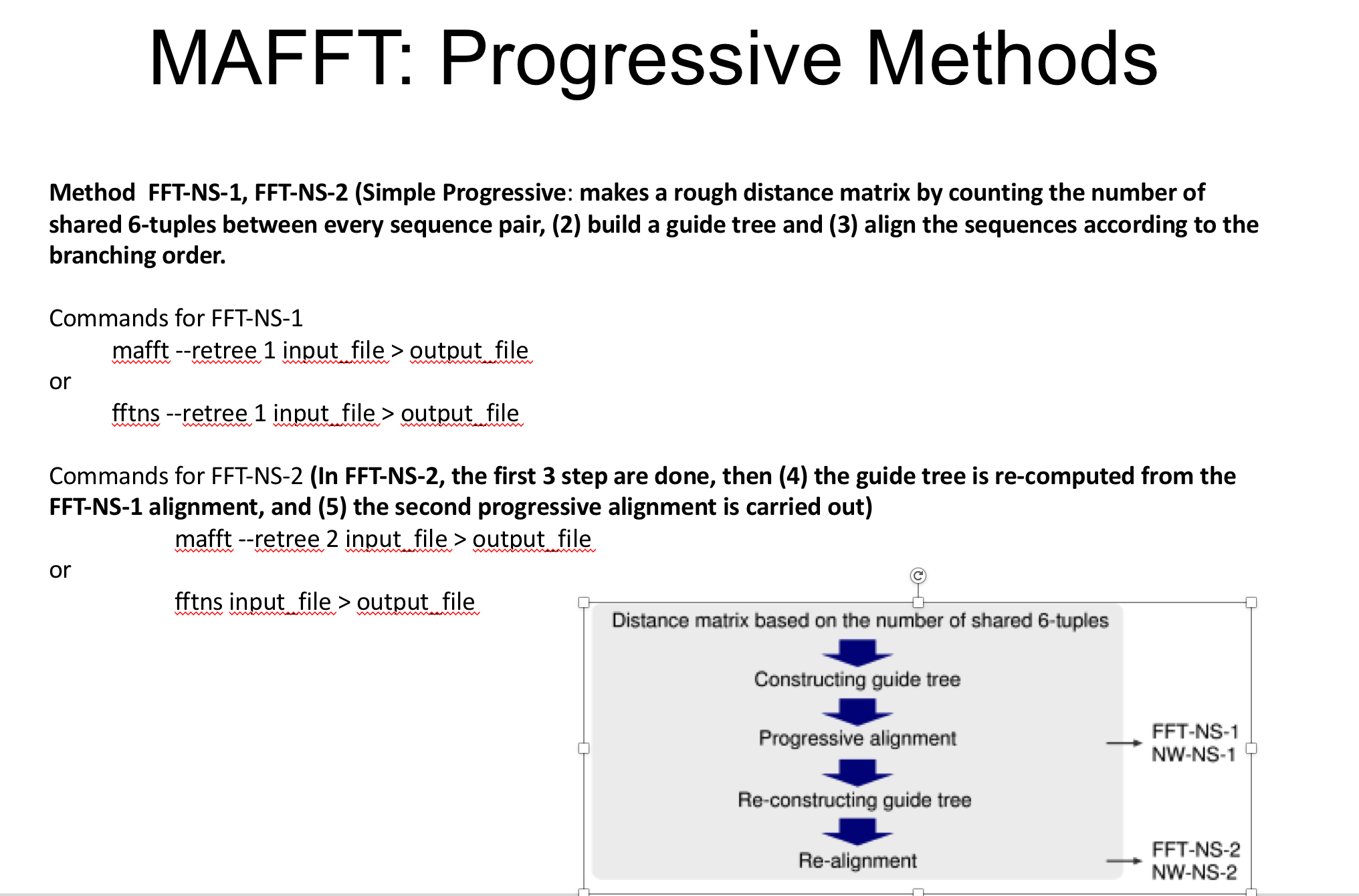
clustalo --p1=Nectin4PrimatesOmega.out --p2=Nectin4CarnivoresOmega.out -o Nectin4PrimatesCarnivoresOmega.out

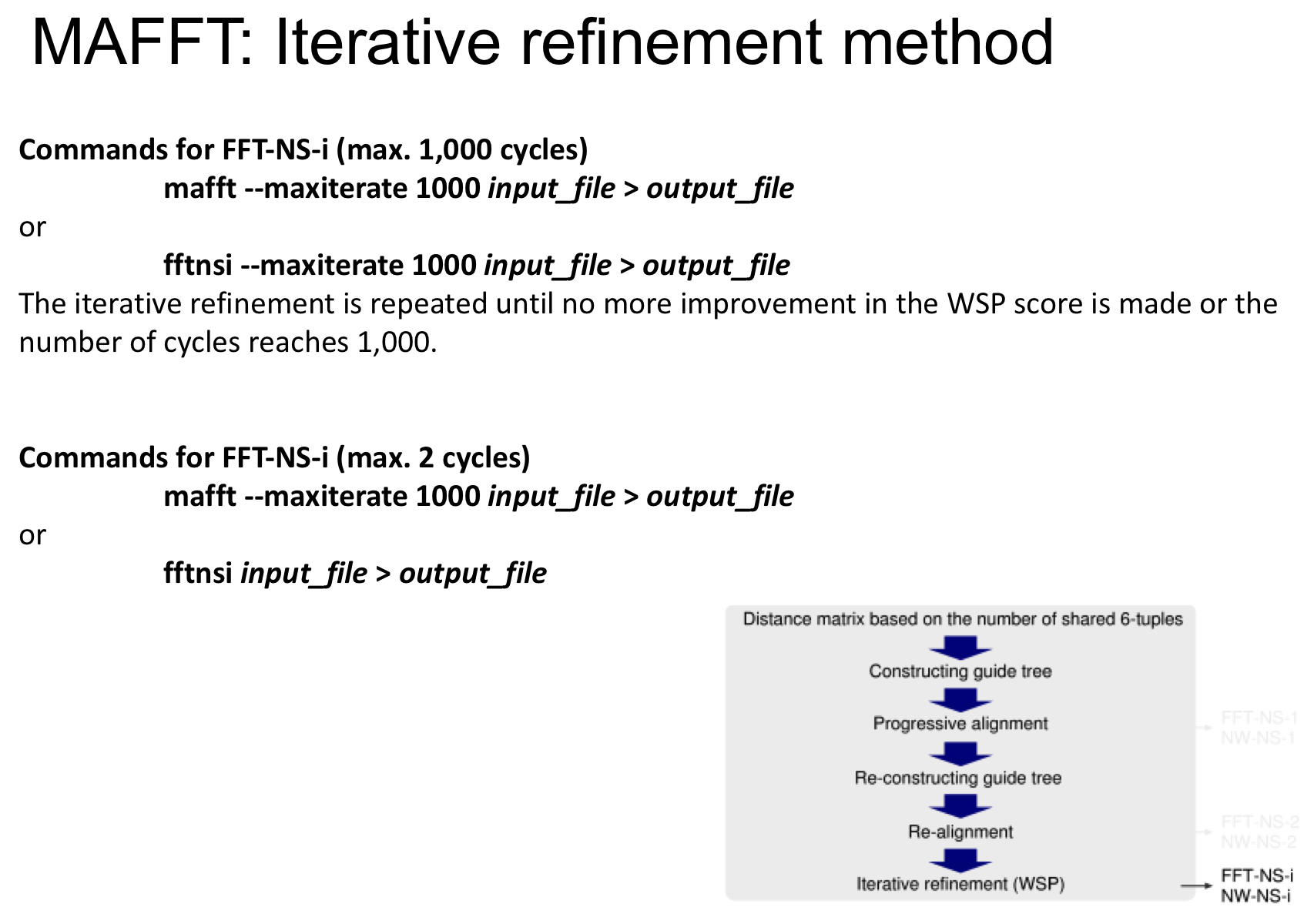
**Step 9.** You can also align new file of unaligned sequences to an existing profile using the command

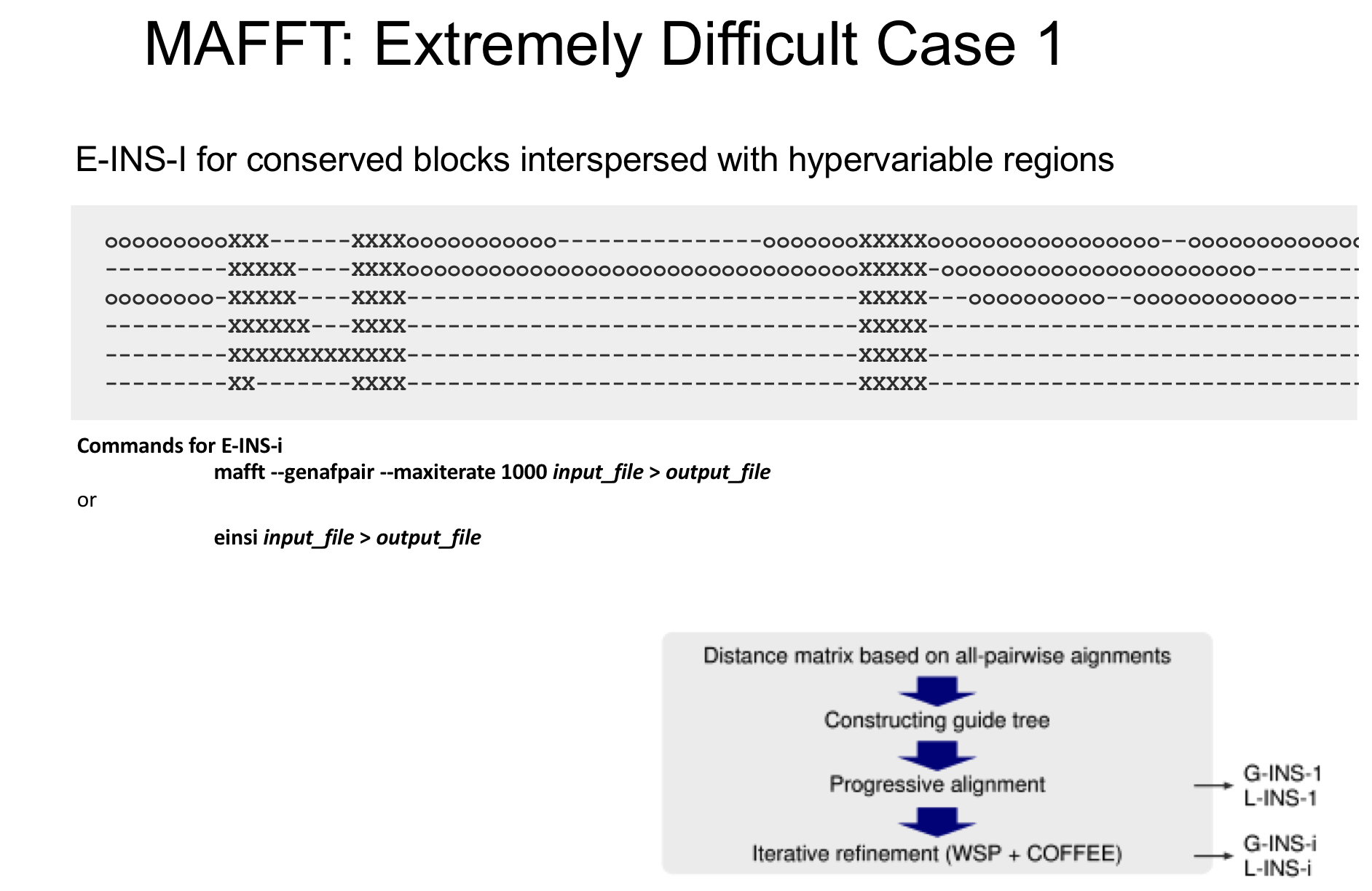
Clustalo --p1=Nectin4PrimatesOmega.out -i Nectin4Carnivores.fasta -o Nectin4PrimatesCarnivoresOmega.out2

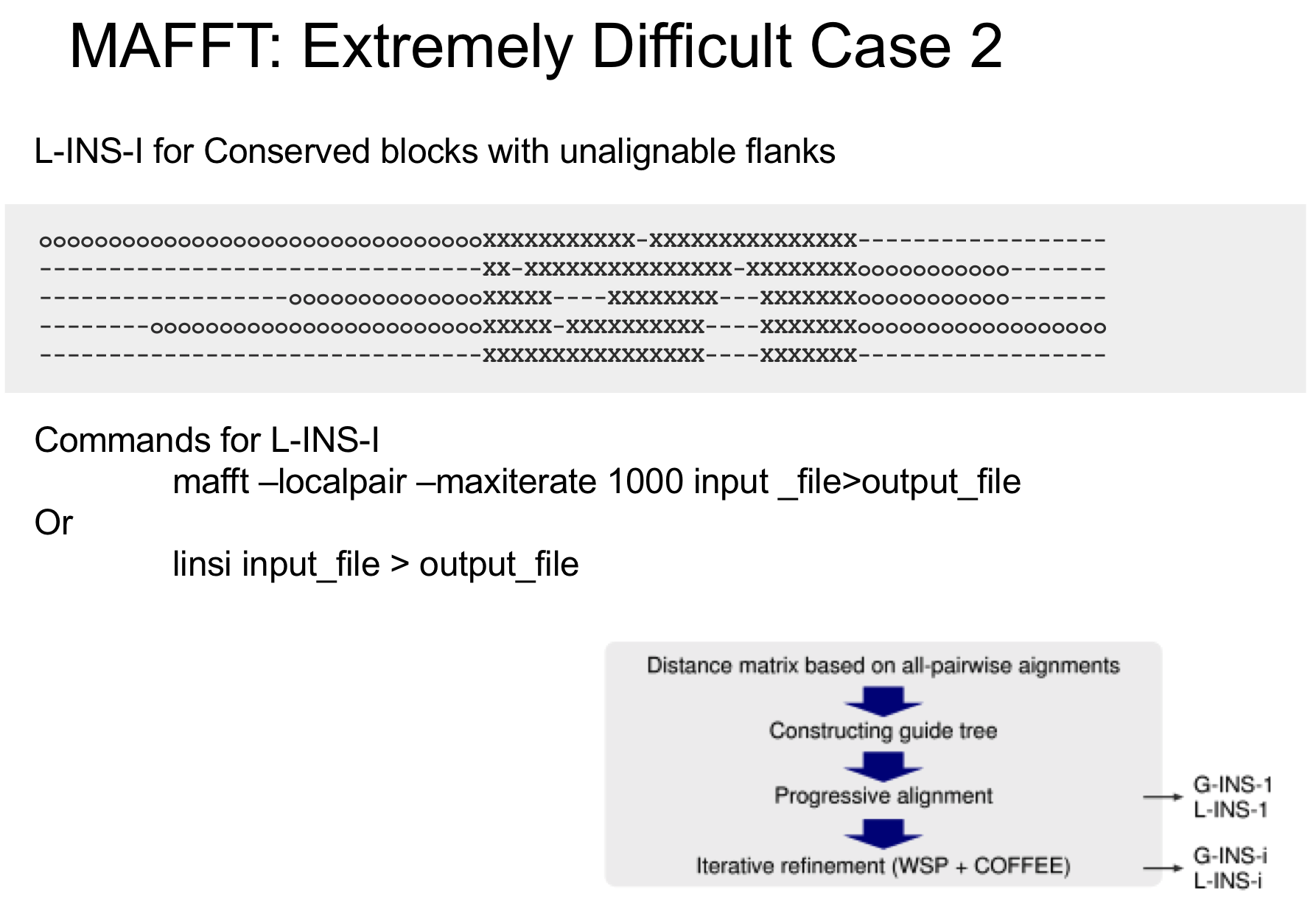
**Step 10.** Open file in **Aliview** to check alignment

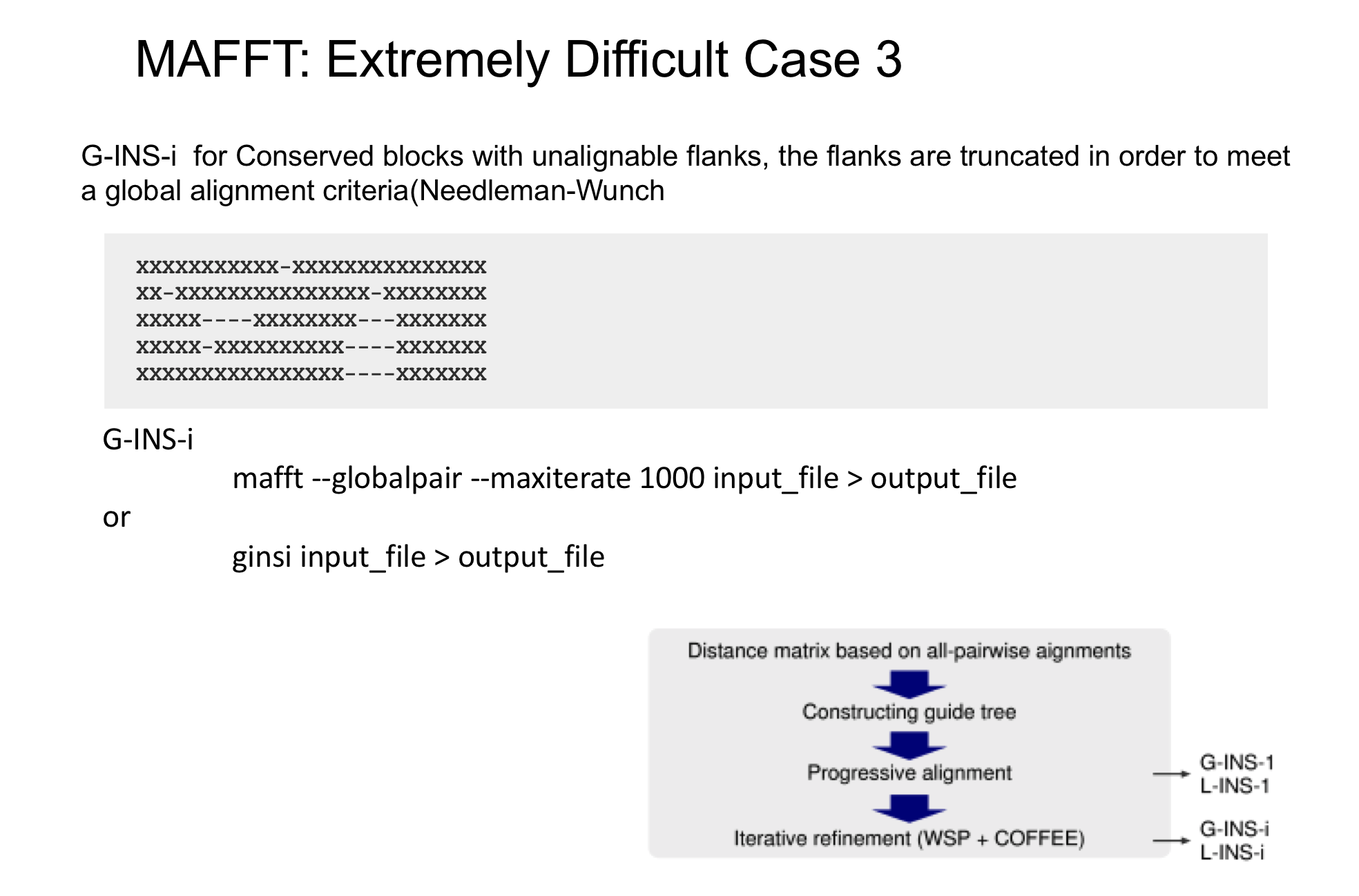
**Part 3. Using MAFFT for Carnivores and Primates Example. MAFFT is fast, powerful with high scalability capable of handling very large data sets. MAFFT Has different functions depending on the types of sequences to be aligned. Here are some commands for different cases.**

****









**Step 1:** Open terminal and go to the Infiles folder

cd /Users/gdw/Desktop/Alignment\_Exercise/Infiles

**Step 2:** Type mafft. An interactive menu appears to help you select options.

Alternatively, you can use command line to specify the options.

**Step 3**: We will use commands

mafft --thread -1 --globalpair --maxiterate 16 --reorder Nectin4Primates.fasta > /Users/gdw/Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesMafft.out

**Step 4**: Open output in **Aliview** using commands from Part 1. View alignment *Nectin4PrimatesMafft.out*.

**Step 5:** Return to terminal and Infiles folder

cd /Users/gdw/Desktop/Alignment\_Exercise/Infiles

**Step 6:** **Mafft** also allows you to add sequences to existing alignments. We can add the carnivores to the alignment using the following commands:

mafft --add Nectin4Carnivores.fasta --reorder Nectin4PrimatesMafft.out >

/Users/gdw/Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresMafft.out

**Step 7:** Open Nectin4PrimatesCarnivoresMafft.out in **Aliview** using commands from Part 1.

**PART 4: PRANK is a command line alignment program that is different from others as it strives to find the most evolutionarily correct alignment. It is focused on the reconstruction of evolutionary homology particularly with respect to placement of indels. However. If sequences are very divergent than will not work correctly.**

**For those interested in web-based application for full visualization- go to WASABI**

**http://wasabiapp.org**

**Publication: https://academic.oup.com/mbe/article/33/4/1126/2579418**

**Step 1:** Open Terminal and go to Infiles.

cd /Users/gdw/Desktop/Alignment\_Exercise/Infiles

Typical command if you have a guide tree.

prank -d=input\_file -t=tree\_file -o=output\_file -F

Here*, input\_file* is the name of the file with input sequences in FASTA format, *tree\_file* is the name of the file with a phylogeny with branch lengths relating these sequences, and *output\_file* is the name of the file (with extension .best.fas) where the resulting alignment will be written in FASTA format; -F specifies that the inference of insertions should be trusted and sites appearing as insertions should not be aligned at the later stages of the process

**Step 1**: Alignment of Primate Sequences. For our example we do not have a guide tree established so will let PRANK create one for Nectin4Primates alignment.

prank -d=Nectin4Primates.fasta -o=/Users/gdw/Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesPrank.out -F

**Step 2:** Open output in Aliview using commands from Part 1. View alignment.

**Step 3:** Return to terminal-

Cd /Users/gdw/Desktop/Alignment\_Exercise/Infiles

**Step 4:** Align Carnivore Sequences.

prank -d=Nectin4Carnivores.fasta -o=/Users/gdw/Desktop/Alignment\_Exercise/Student\_files/Nectin4CarnivoresPrank.out -F

**Step 5**: Combine the two PRANK alignments-each file should be aligned previously.

Prank -d1= Nectin4PrimatesPrank.out -d2= Nectin4CarnivoresPrank.out -o=/Users/gdw/ /Desktop/Alignment\_Exercise/Student\_files/Nectin4PrimatesCarnivoresPrank.out

**Step 8: Compare Alignments between Muscle, ClustalX (omega), Mafft, PRANK by opening all 4 files in Aliview. Are they identical? Which would you use?**

**Part 4: If we have time, repeat these exercises using** **Nectin4Cetaceans.fasta** and **Nectin4Ruminants.fasta**